

## Effect of Training on Exposure Judgment Accuracy of Industrial Hygienists

Monika Vadali , Gurumurthy Ramachandran , John R Mulhausen & Sudipto Banerjee

To cite this article: Monika Vadali , Gurumurthy Ramachandran , John R Mulhausen & Sudipto Banerjee (2012) Effect of Training on Exposure Judgment Accuracy of Industrial Hygienists, Journal of Occupational and Environmental Hygiene, 9:4, 242-256, DOI: [10.1080/15459624.2012.666470](https://doi.org/10.1080/15459624.2012.666470)

To link to this article: <https://doi.org/10.1080/15459624.2012.666470>



Accepted author version posted online: 21 Feb 2012.  
Published online: 30 Mar 2012.



Submit your article to this journal [↗](#)



Article views: 675



View related articles [↗](#)



Citing articles: 7 View citing articles [↗](#)

# Effect of Training on Exposure Judgment Accuracy of Industrial Hygienists

Monika Vadali,<sup>1</sup> Gurumurthy Ramachandran,<sup>1</sup> John R Mulhausen,<sup>2</sup> and Sudipto Banerjee<sup>3</sup>

<sup>1</sup>Division of Environmental Health Sciences, School of Public Health, University of Minnesota, Minneapolis, Minnesota

<sup>2</sup>3M Company, St. Paul, Minnesota

<sup>3</sup>Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, Minnesota

*Results are presented from a study that investigated the effect of data interpretation training on exposure judgment accuracy of industrial hygienists across several companies in different industry sectors. Participating companies provided monitoring information on specific exposure tasks. Forty-nine hygienists from six companies participated in the study, and 22 industrial tasks were evaluated. The number of monitoring data points for individual tasks varied between 5 and 24. After reviewing all available basic characterization information for the job, task, and chemical, hygienists were asked to provide their judgment on the probability of the 95th percentile of the underlying exposure distribution being located in one of four exposure categories relative to the occupational exposure limit as outlined in the AIHA® exposure assessment strategy. Ninety-three qualitative judgments (i.e., without reviewing monitoring data) and 2142 quantitative judgments (i.e., those made after reviewing monitoring data) were obtained. Data interpretation training, with simple rules of thumb for estimating 95th percentiles, was provided to all hygienists. A data interpretation test was administered before and after training. All exposure task judgments were collected before and after training. Data interpretation test accuracy for the hygienists increased from 48% to 67% after training ( $p < 0.001$ ) and a significant underestimation bias was removed. Hygienist quantitative task judgment accuracy improved from 46% to 69% ( $p < 0.001$ ) post-training. Accuracy results showed good improvement in industrial hygienists' quantitative judgments as a result of training. Hence, the use of statistical tools is promoted to improve judgments based on monitoring data and provide feedback and calibration to improve qualitative judgments. It may be worthwhile to develop standard training programs to improve exposure judgments.*

**Keywords** bias in exposure judgment, data interpretation exposure judgment accuracy, professional judgment

Correspondence to: Gurumurthy Ramachandran, University of Minnesota, Division of Environmental Health Sciences, School of Public Health, MMC 807, 420 Delaware St. SE, Minneapolis, MN 55455; e-mail: ramac002@umn.edu.

## INTRODUCTION

Most exposure assessment strategies require the work force to be categorized into similar exposure groups, or SEGs.<sup>(1–3)</sup> After categorizing the work force into SEGs, the industrial hygienist uses a combination of professional judgment, personal experience with a given type of operation, review of exposures from similar operations, and/or exposure predictions developed using physical/chemical or empirical exposure modeling techniques to assign an initial “exposure rating” and prioritize their SEGs for further actions. Those actions may include quantitative monitoring or various types of control measures. Based on this prioritization, a baseline monitoring campaign is carried out for some SEGs, and the measurement data collected are used to refine the initial rating and determine if the exposure distribution is acceptable. Acceptability is commonly evaluated by comparing an upper percentile such as the true group 95th percentile to the occupational exposure limit (OEL). Further, the degree of acceptability can be described by how far the decision statistic is from the OEL, and thus, the SEG's exposure can be classified into one of four categories: (1) highly controlled, (2) well controlled, (3) controlled, or (4) poorly controlled (Table I).

The process of estimating the probability of the decision statistic (the 95th percentile of the OEL) being located in each of the four exposure categories (highly controlled to poorly controlled as described above) may use subjective professional judgment, mathematical modeling, monitoring data, or some combination of these. The estimate of a set of such probabilities will be referred to as a *judgment* in this article.

Although the AIHA® strategy suggests making 6 to 10 exposure measurements in a SEG,<sup>(2,3)</sup> exposure judgments are often based on fewer measurements. In many situations, exposure assessment may be required for several chemical

**TABLE I. Exposure Category Rating Scheme.**

| AIHA Exposure Rating | Proposed Control Zone Description | Qualitative Description  | AIHA-Recommended Statistical Interpretation |
|----------------------|-----------------------------------|--|---|
| 1                    | Highly controlled                 | Exposures infrequently exceed 10% of limit                             | 95th Percentile <0.10*OEL                   |
| 2                    | Well controlled                   | Exposures infrequently exceed 50% of limit and rarely exceed the limit | 0.10*OEL < 95th Percentile <0.5*OEL         |
| 3                    | Controlled                        | Exposures infrequently exceed the limit                                | 0.5*OEL < 95th Percentile <OEL              |
| 4                    | Poorly controlled                 | Exposures frequently exceed the limit                                  | OEL < 95th Percentile                       |

species simultaneously. Many facilities have hundreds of process/task/substance combinations, making a complete quantitative exposure assessment for each all but impossible to accomplish. Anecdotally, it is estimated that greater than 90% of exposure ratings may be based on professional judgments without any monitoring data. Thus, there is a heavy reliance on the accuracy of professional judgments and the ability of industrial hygienists to correctly integrate them with monitoring data to reach an accurate exposure determination. Professional judgment is the aptitude of an experienced professional to make correct inferences using incomplete data.<sup>(3)</sup> It is based on factors such as observation of the workplace, analogy, previous knowledge, and experience. The aim of this research was to study accuracy of exposure judgments of industrial hygienists from a range of companies and to investigate the effect of targeted training on accuracy.

Several studies have evaluated the ability of industrial hygienists to assess exposures accurately.<sup>(4-17)</sup> Hawkins and Evans<sup>(4)</sup> evaluated the professional judgments of 24 industrial hygienists by asking them to estimate toluene exposures for a single process. Industrial hygienists provided estimates of the mean, median, range, and 90th percentile exposures using only qualitative information at first, and then using limited historical exposure data. Their estimates were then compared to a distribution determined from 134 exposure measurements from the process. Industrial hygienists were better at estimating the 90th percentile than the median, and their judgment improved significantly when presented with historical data.

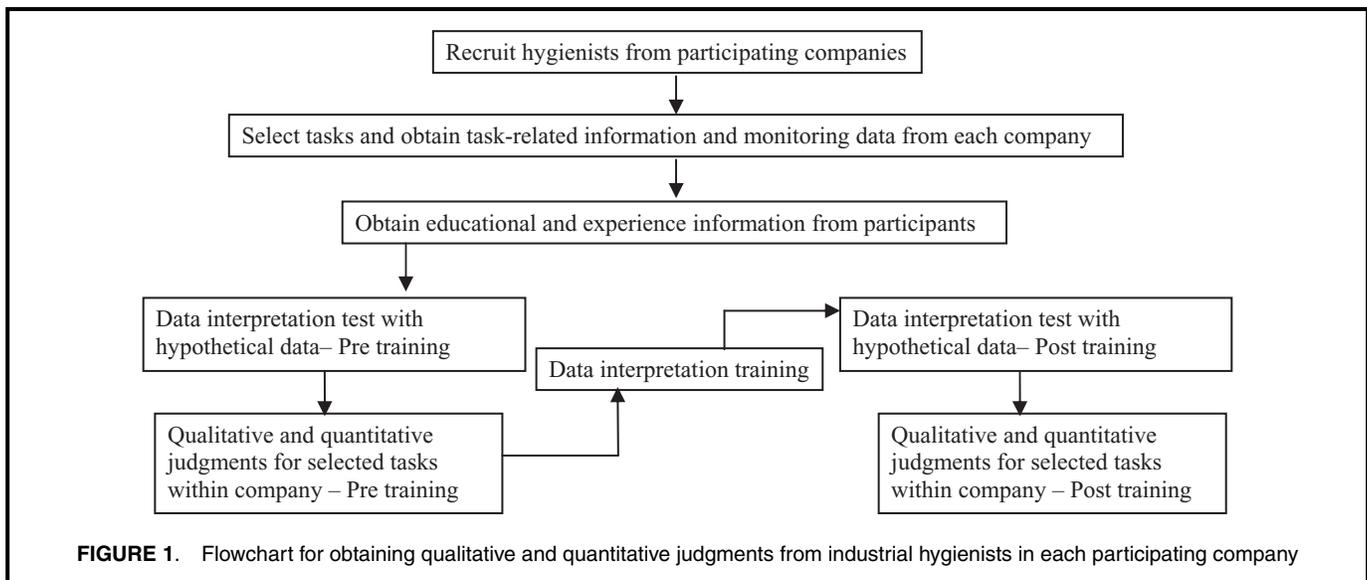
Walker et al.<sup>(5,6)</sup> evaluated the use of expert judgment to estimate non-occupational exposures to benzene. Accurate, consistent judgments could be obtained from the experts. When the exposure estimates were compared to the actual measurements, it was found that the experts were better at estimating 90th percentiles than the mean. Logan et al.<sup>(7)</sup> investigated the effect of exposure data interpretation training on accuracy of exposure judgments in a group of industrial hygienists. The 75 industrial hygienists were shown videos of industrial tasks and associated monitoring data, and their exposure judgments were collected before and after training for interpretation of small data sets. The percentage of correct judgments for exposure judgments based on monitoring data increased from 43% pre-training to 63% post-training. Kromhout et al.<sup>(10)</sup> compared the exposure assessments of industrial hygienists,

task supervisors, and workers on 97 tasks and found that there was a significant correlation between subjective ratings based on a four-point scale and the measured mean exposures. A strategy combining the use of subjective judgments and exposure measurements was hypothesized to lead to better exposure estimates.

Macaluso et al.<sup>(8)</sup> looked at 5 experts judging 695 different department-year-job combinations from 29 work histories to assess solvent exposures in the painting operation of a car assembly plant. It was observed that for low-exposure categories there was considerable disagreement between the experts. De Cock et al.<sup>(9)</sup> analyzed rankings of dermal and respiratory exposures in 14 fruit growing pesticide application tasks provided by 15 experts. A statistically significant difference was found between the exposure ratings and actual measurements. Clear differences, based on level of expertise, were found between experts' exposure ratings. Accurate exposure rating was found to correlate with the level of expertise. Stewart et al.<sup>(11)</sup> looked at three industrial hygienists evaluating 300 jobs for formaldehyde exposures. Industrial hygienists were given six cycles of increasing information regarding the exposures. It was found that the agreement between the industrial hygienists was slightly greater than that expected due to random chance. There was moderate improvement in overall agreement over the cycles, measured by the weighted kappa statistic. Other studies of professional judgment in the context of exposure assessment showed that in most scenarios there was poor agreement between the experts and that agreement decreased when the exposure levels were very low.<sup>(12-17)</sup> This was the case even when actual measurements were available.

In summary, experts appear to be fairly adept at estimating upper percentiles as well as arithmetic means of exposure distributions. Accuracy tends to be correlated with the level of expertise. Judgments become more accurate with more available data. It has also been suggested that the use of simple statistical rules of thumb improves accuracy.<sup>(7)</sup>

This study looks at the accuracy of exposure judgments of industrial hygienists for several task-related exposures. Industrial hygienists were recruited from six companies that agreed to participate in the study. Their exposure judgments were collected via online conference call. The judgments collected were analyzed for accuracy. The effect of data interpretation training on accuracy was also studied.



## METHODS

To recruit a wide range of industrial hygienists, presentations were made at the Professional Conference on Industrial Hygiene (PCIH) in 2005 and at the American Industrial Hygiene Conference and Expo (AIHCE) in 2006. Interested companies were contacted and invited to participate. Figure 1 illustrates the procedure used for recruiting industrial hygienists, task selection, and the sequence used for data collection and training.

Within each participating company, one person served as our principal contact. This person recruited the industrial hygienists who participated from that company, selected the tasks for which monitoring data were collected, and provided all the necessary task process details to the participating hygienists. Companies were asked to choose tasks such that their exposures spanned the range of the four AIHA exposure categories. Each task was clearly defined with details of what the worker does on the job. Each company was asked to provide a minimum of 10 hygienist-task combinations. Each task was to have 8–10 personal exposure monitoring data. A total of 22 tasks from all six companies were evaluated.

The hygienists were recruited based solely on their availability and provided informed consent to participate in this exercise. The study had a total of 49 industrial hygienists from six companies. Hygienists from each company provided judgments for tasks selected only from their company. Table II shows the summary statistics of the educational and experience backgrounds of all participating hygienists. There is a wide variation, between 4 months to 26 years, in the experience that participants had in their current job. Of the participants, 84% had a master's degree in industrial hygiene, environmental health, or a related science field. Slightly more than half the industrial hygienists in the study group had some certification, with 48% having either a CIH certification or both CIH and CSP certification. While 39% of industrial hygienists had more than 10 years of experience making exposure judgments, 94%

of them had engaged in exposure assessment activities within the past year. All industrial hygienists had some level of experience documenting tasks using the AIHA strategy and some knowledge of statistics, with 27% of them familiar with calculating confidence intervals or tolerance limits for 95th percentiles.

Each company provided monitoring data for the specific tasks that were chosen for judgment elicitation. The industrial hygienists were asked to review all available information for each task that they evaluated, including the workplace process, chemicals used in the task, material safety data sheets, controls, personal protective equipment worn by the workers, duration of the task, and task environment parameters, such as the general room ventilation. The hygienists were also encouraged to personally view the actual facility and talk to plant personnel regarding the tasks.

The hygienists were asked to give their initial judgment based on their knowledge of the task alone, without looking at any monitoring data. The first monitoring data point was then revealed to them, and they were allowed to modify their initial judgment on the basis of this new information. Subsequent data points were revealed, one at a time, and the hygienists were allowed to modify their judgment after each data point.

To assess the accuracy of industrial hygienists' judgments, the true exposure category, or "reference," was defined for each task based solely on the monitoring data provided by the company. The true exposure category is defined as the category most likely to contain the true 95th percentile exposure based on the complete set of monitoring data.<sup>(18)</sup> A total of 22 tasks were evaluated (Table III). The number of monitoring data points for each task varied between 5 and 24. While the companies had been requested to provide tasks with exposures in all four categories, 82% of the tasks were Category 4 exposures (as evaluated by the complete set of monitoring data for each task). This is an important consideration while interpreting the results.

**TABLE II. Summary Data for Educational and Exposure Assessment-Related Professional Experience for All Participants (N = 49)**

|  |  |   |  |   |  |
|--|--|---|--|---|--|
| Years in current job   | <1<br>2%   | 1 to 5<br>34%   | 6 to 10<br>28%   | >10<br>36%  |  |
| Highest educational degree   | Associate degree in<br>EHS or science<br>2%                | BS degree in EHS<br>or science<br>14%                             | MS degree in EHS<br>or science<br>84%  | Doctoral degree in<br>EHS or science<br>0%  |  |
| Certifications   | CIH<br>24%   | CSP<br>2%   | CIH and CSP<br>24%   | IHIT<br>2%  | None<br>47%                            |
| Number of years making<br>exposure judgments                               | <1 year<br>6%  | 1 to 3<br>16%   | 3 to 5<br>14%  | 5 to 10<br>24%  | >10<br>39%                             |
| Number of years since<br>actively engaged in doing<br>exposure assessments | <1 year<br>94%   | 1 to 3<br>4%  | 3 to 5<br>2%   | 5 to 10<br>0%   | >10<br>0%                              |
| Estimate of total number of<br>job tasks documented<br>using AIHA strategy | <10<br>20%   | 10 to 100<br>31%  | 100 to 1000<br>24%   | 1000 to 5000<br>16%   | >5000<br>8%                            |
| Career air sampling<br>experience  | <3 Air sampling<br>surveys<br>0%                           | 3 to 20 Air<br>sampling surveys<br>8%                             | 20 to 100 Air<br>sampling surveys<br>29%   | 100 to 200 Air<br>sampling surveys<br>29%   | >200 Air<br>sampling<br>surveys<br>35% |
| Statistical experience   | No understanding<br>of normal and<br>lognormal stats<br>0% | Can do basic<br>lognormal stat<br>calculations<br>(GM,GSD)<br>41% | Familiar with<br>calculating point<br>estimate of 95th<br>percentiles or<br>exceedance<br>fractions<br>33% | Familiar with<br>calculating<br>confidence<br>intervals or<br>tolerance limits for<br>95th percentiles<br>27% |  |

We collected 93 qualitative task judgments (i.e., without reviewing monitoring data) and 2142 quantitative task judgments (i.e., those made after reviewing monitoring data) from the industrial hygienists before and after training. The judgments elicited were in the form of the probability of the 95th percentile being in each of the four AIHA categories. This can be represented graphically, as a set of decision probabilities (Figure 2).

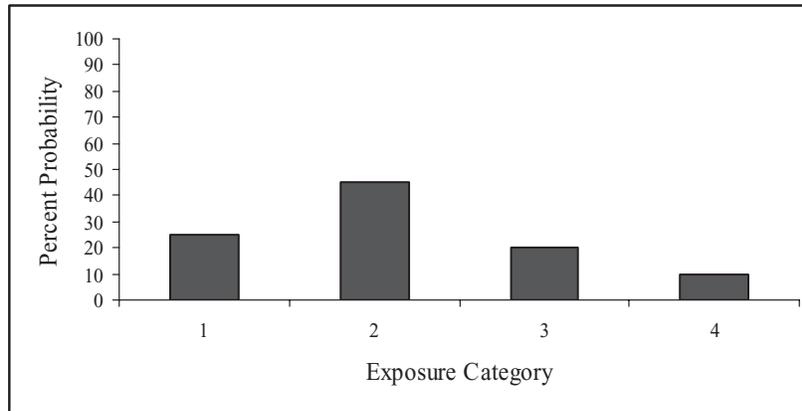
The probabilities were provided on a scale of 1–100 in whole numbers for each of the four categories. No category could have a probability of zero (for mathematical convenience), and one of the four categories had to have the highest probability. This reflects the category that the industrial hygienist thinks is the most probable one in which the 95th percentile of the exposure distribution is located for that particular task. Further control decisions are made on the basis of this assessment. Thus, the minimum probability that can be assigned to a category is 1%, and the maximum that can be assigned to a category is 97%.

Data interpretation training was provided to the participants after the first round of judgments were made for each task. This training covered basic statistics and properties of lognormal distributions. Study participants were taught some

simple rules of thumb to estimate the 95th percentile rapidly for small data sets. The industrial hygienists were given data interpretation tests (DIT) using small simulated monitoring data sets before and after training (Table IV). The DIT had eight data sets with the number of simulated data points varying from 1–8. The OEL for all data sets was assumed to be 100. Hygienists were asked to predict the probabilities of the 95th percentile being located in each of the four exposure categories for each data set. For example, the third data set in Table IV has four data points 15, 12, 23, and 9 ppm. The industrial hygienists would enter their judgments for that data set in the Data Set #3 column. There were a total of five available DITs out of which two were assigned randomly to each industrial hygienist. The entire exercise was performed again, post-training, for all industrial hygienists with the original tasks and a new DIT consisting of eight different data sets.

Based on the judgment data gathered, the following null hypotheses were tested:

1. Data interpretation training has no effect on the industrial hygienists' data interpretation test judgment accuracy.



**FIGURE 2.** Example of a decision chart showing the probability of the 95th percentile being in each of the four categories

2. Data interpretation training has no effect on industrial hygienists' task judgment accuracy.

Two definitions were used for judgment accuracy. The first definition, called absolute accuracy, is based on how closely the hygienist's assignment of probabilities in the four categories

matches the probabilities based on the complete set of monitoring data (the reference judgment). For the data interpretation tests and the task judgments, the absolute accuracy for each industrial hygienist was calculated using the formula:

$$Absolute\ accuracy = 1 - \frac{\sqrt{\sum_{i=1}^4 (Tr_i - IH_i)^2}}{(96\sqrt{2} - 0)} \quad (1)$$

**TABLE III.** Task Monitoring Data Samples and Reference Judgment Category for Each Task

| Task | Number of Samples Used for Making Judgments and for Calculation of Reference Judgment | Most Probable Exposure Category Calculated for Each Task (Reference Judgment) |
|------|---|---|
| 1    | 8   | 4   |
| 2    | 7   | 4   |
| 3    | 7   | 4   |
| 4    | 6   | 4   |
| 5    | 7   | 4   |
| 6    | 10  | 2   |
| 7    | 17  | 4   |
| 8    | 7   | 4   |
| 9    | 5   | 4   |
| 10   | 8   | 4   |
| 11   | 24  | 4   |
| 12   | 6   | 4   |
| 13   | 6   | 4   |
| 14   | 7   | 4   |
| 15   | 11  | 4   |
| 16   | 10  | 4   |
| 17   | 8   | 4   |
| 18   | 7   | 4   |
| 19   | 8   | 4   |
| 20   | 8   | 3   |
| 21   | 11  | 3   |
| 22   | 8   | 2   |

where

$Tr_i$  = True probability assignment for the  $i$ th category based on complete monitoring data set (reference category)

$IH_i$  = Industrial hygienist's probability assessment for  $i$ th category

$96\sqrt{2}$  = maximum possible difference between an industrial hygienist's judgment and the reference category

0 = minimum possible difference between an industrial hygienist's judgment and the reference category

The sum of all four categories should add up to 100; the minimum probability in any category cannot be below 1% (for mathematical convenience in the algorithm), and the maximum probability in any one category cannot exceed 97. For example, if the reference judgment is a Category 4 exposure with 97% probability, but the industrial hygienist predicts Category 1 with 97% probability, then the numerator of Eq. 1 would represent the maximum difference, which equals  $96\sqrt{2}$ . In this case, absolute accuracy is 0. The minimum difference between any industrial hygienist judgment and the reference is 0, in which case the absolute accuracy is 1. Absolute accuracy accounts for not just whether the hygienist's most likely category matched the category determined by the monitoring data but also the difference in the spread of probabilities in the four categories between the hygienist's judgment and that based on monitoring data.

The second definition, called categorical accuracy, was based on the deviation from the reference (or true) category. Categorical accuracy was also calculated for both data interpretation tests as well as task data judgments. If the industrial hygienist predicted the same category as the reference exposure category then the categorical accuracy was assigned a value

**TABLE IV. Sample Data Interpretation Test**

| STATISTICAL TEST DATA SHEET  |             |             |             |             |             |             |             |             |
|--|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Name of Company  |             |             |             |             |             |             |             |             |
| Hygienists ID  |             |             |             |             |             |             |             |             |
| OEL for all Data 100   |             |             |             |             |             |             |             |             |
| Sets   |             |             |             |             |             |             |             |             |
|  | Sample      |
|  | Data Set #1 | Data Set #2 | Data Set #3 | Data Set #4 | Data Set #5 | Data Set #6 | Data Set #7 | Data Set #8 |
|  | 7           | 55          | 15          | 4           | 1           | 23          | 112         | 1           |
|  | 17          |             | 12          |             | 1           | 41          | 86          | 2           |
|  | 20          |             | 23          |             | 8           | 12          | 72          | 10          |
|  | 24          |             | 9           |             | 1           | 8           |             | 1           |
|  | 37          |             |             |             |             | 18          |             | 4           |
|  | 40          |             |             |             |             | 36          |             |             |
| Make your judgments on the above Statistics Test Data in the following columns |             |             |             |             |             |             |             |             |
|  | Data Set #1 | Data Set #2 | Data Set #3 | Data Set #4 | Data Set #5 | Data Set #6 | Data Set #7 | Data Set #8 |
| <10% OEL   |             |             | 5           |             |             |             |             |             |
| 10–50% OEL   |             |             | 60          |             |             |             |             |             |
| 50–100% OEL  |             |             | 30          |             |             |             |             |             |
| >100% OEL  |             |             | 5           |             |             |             |             |             |
| Check  | 0           | 0           | 100         | 0           | 0           | 0           | 0           | 0           |
| Have you ever taken this statistical test before?                              |             |             |             | Yes         | No          |             |             |             |
| If yes, Where?   |             |             |             |             |             |             |             |             |
| If yes, how many times?  |             |             |             |             |             |             |             |             |

of 1. If the industrial hygienist predicted a category that was higher than the reference (could be 1, 2, or 3 categories higher) the categorical accuracy was assigned a value of 2 for that judgment. If the industrial hygienist predicted a category lower than the reference (could be 1, 2, or 3 categories lower), the categorical accuracy was assigned a value of 3. The assignment of numbers 1, 2, and 3 for categorical accuracy are only nominal values and not ordinal. They are not ranked in any way, but the numbers are assigned only to identify the category over- or underprediction.

1. correct prediction
2. overprediction
3. underprediction

Accuracy for task judgments was calculated using two other metrics:

$$\begin{aligned}
 & \text{Fraction of correct decisions by } j^{\text{th}} \text{ hygienist (F}_j\text{)} \\
 &= \frac{\text{Number of correct decisions by IH}_j}{\text{Total number of scenarios (N) looked at by IH}_j} \quad (2)
 \end{aligned}$$

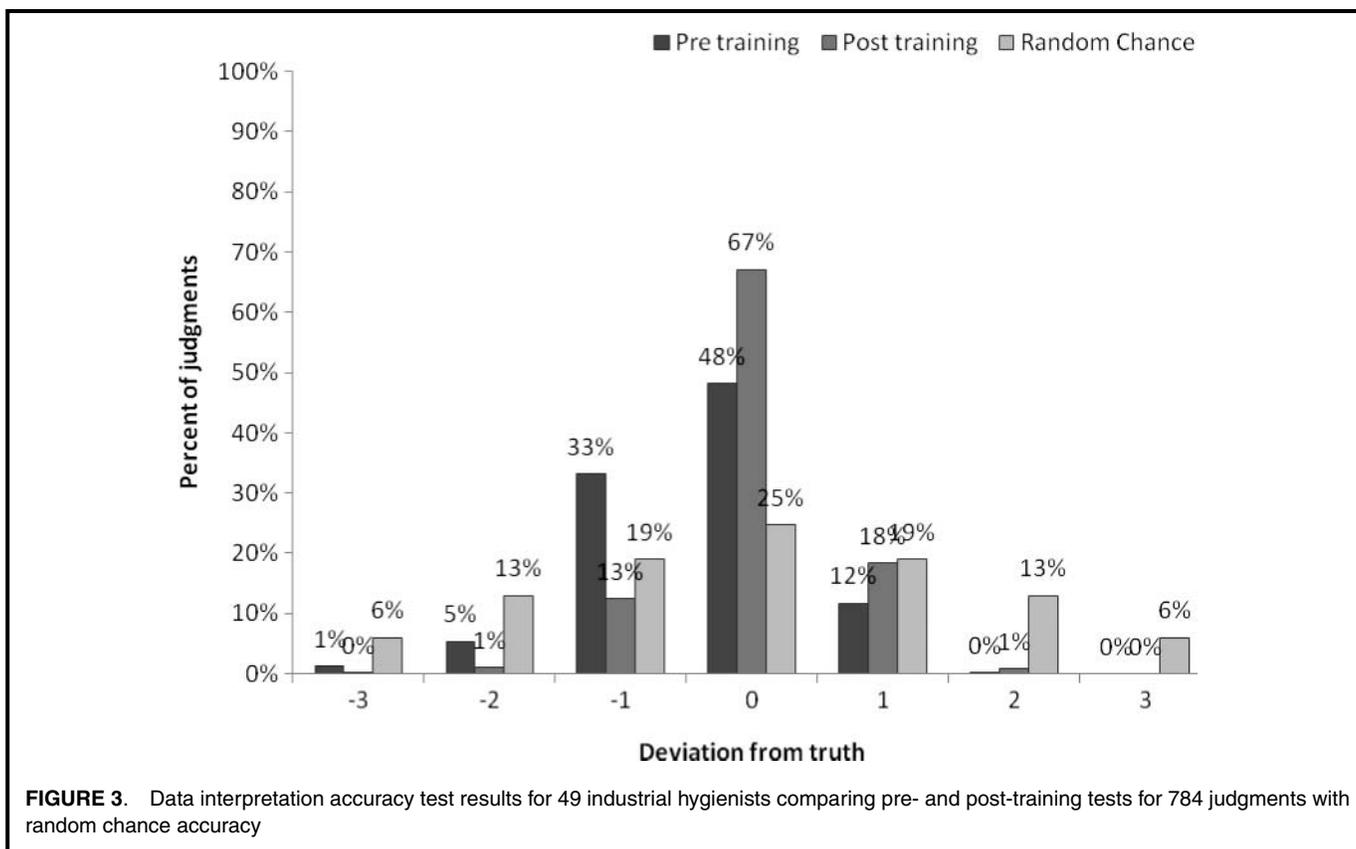
$$\begin{aligned}
 &= \frac{\text{Fraction of IHs making correct decisions}}{\text{Number of IHs making correct decisions}} \\
 &= \frac{\text{Number of IHs making correct decisions}}{\text{Total number of IHs}} \quad (3)
 \end{aligned}$$

## RESULTS AND DISCUSSION

Data interpretation test results are discussed first, followed by the actual task judgment results. For the data interpretation results, the bias in judgments before and after training is discussed, followed by the discussion of statistical modeling of absolute accuracy and categorical accuracy. For task judgments, a discussion of bias in judgments before and after training is followed by a brief discussion of pre-training qualitative judgments. Fraction of correct judgments by industrial hygienists and a discussion of fraction of industrial hygienists making correct judgments are followed by the absolute accuracy and categorical accuracy models for task judgments.

### Data Interpretation Test Results

The DITs were designed to study how accurately industrial hygienists interpret small data sets to predict the true 95th percentile exposure category. Each DIT had eight data sets, and each industrial hygienist provided judgments for two such tests, for a total of 784 judgments from 49 industrial hygienists. The DIT judgments were investigated to study how well industrial hygienists predict the correct exposure category as well as the deviation of judgments from the correct category. Since there are four exposure categories, a maximum of overprediction or underprediction by three categories is possible. Figure 3 shows that before training, almost 40% of



DIT judgments made by industrial hygienists underestimated the true exposure category, and only 48% of judgments were able to predict the correct exposure category. This could be due to a bias in underestimating upper percentiles of skewed, lognormal distributions, and incorrectly visualizing the data as being normally distributed instead. Monitoring data come from underlying lognormal distributions that have long upper tails that need to be taken into account when predicting probability. For small data sets, when quick decisions are made without statistical analyses, this upper tail may be underestimated, resulting in underestimation bias in decisions.

Industrial hygienists did better than random chance even before the training was provided. If hygienists made exposure judgments completely randomly, each exposure category has an equal probability of being selected, and the correct category will be selected 25% of the time. The over- or underprediction probabilities can also be calculated by using a random number generator to select a number between 1 and 4 for each of the 784 DIT tests. The deviation of the judgments made by random chance from the correct judgment for the DIT was calculated and plotted as bar graphs along with the deviation of the pre- and post-training judgments for DIT as well as task judgments in Figure 3.

The total number of correct judgments made by industrial hygienists on the DITs increased significantly from 48% to 67% ( $p = 0.005$ ) after the training. Post-training results show that more judgments are on the side of overestimation

than underestimation. Compared to the reference, 39% of judgments were underestimated before training, which decreased to 14% after training. Overprediction increased from 12% before training to 19% after training. This indicates a significant effect of training on the accuracy of DIT judgments.

To test Hypothesis 1, a general linear model was fit using PROC GLM (version 9.1; SAS Inc., Cary, N.C.) with absolute accuracy for the DITs as the dependent variable and company (set of dummy variables representing the six companies), training (binary variable, 0 = pre-training and 1 = post-training), and their interactions with industrial hygienist (a set of dummy variables representing the 42 industrial hygienists) as the independent variables. The study was designed to investigate the effect of training on accuracy.

The industrial hygienist's company was also known before the judgment elicitation exercise. Hence, the effect of training and company were study design variables that were investigated to study their effect on accuracy of judgments. The hygienist variable is nested within the company variable (Hygienist\*company). Since individuals are different in the way they perceive information, the industrial hygienists' interactions were also investigated to see if they were significant. Hygienist\*training was the other interaction of interest. The model was found to be significant with an  $r$  value of 0.35. The low value might indicate that there may be other as yet unknown variables that could have explanatory value in

**TABLE V. Results from GLM Model with Data Interpretation Test Absolute Accuracy as the Dependent Variable**

| Variable           | F Value | Pr > F |
|--------------------|---------|--------|
| Company            | 4.28    | 0.0008 |
| Training           | 12.89   | 0.0004 |
| Hygienist*Company  | 4.60    | <.0001 |
| Hygienist*Training | 2.06    | <.0001 |

explaining the variability in the accuracy scores. Table V shows the significant results.

Training was found to be a significant predictor of absolute accuracy ( $p = 0.0004$ ), indicating that the rules of thumb training enabled hygienists to better predict exposure categories. The training interaction with IH was also found to be very significant ( $p < 0.0001$ ), indicating an individual effect whereby individuals have different aptitudes and abilities to assimilate training.

A significant company effect was also observed for absolute accuracy ( $p = 0.0008$ ). While company-specific information was not collected in the study to explain the between-company effect or control for potential confounders, it indicates that the company an industrial hygienist works for may have an effect on decision-making abilities. This could be because of the training provided in the company or the exposure assessment strategy in place at the company. If a company has an exposure assessment strategy similar to the AIHA strategy in place, and if industrial hygienists are routinely trained in making decisions using this strategy, they may perform better than industrial hygienists from companies that do not have any such strategy in place. It could also be that the company variable is a surrogate for other factors, such as whether or not most of the industrial hygienists in a company are certified industrial hygienists (CIHs). In addition, the company interaction with the industrial hygienist was also significant ( $p < 0.0001$ ), indicating a significant within-company effect. This might indicate that there are significant differences between how individual industrial hygienists assimilate the training, strategy, and culture within a company.

To test Hypothesis 1 in a slightly different manner, a cumulative logit model was also fit using PROC LOGISTIC<sup>(19)</sup> with data interpretation test categorical accuracy as the dependent variable and training (binary variable, 0 = pre-training and 1 = post-training), company (set of dummy variables representing the six companies) and their interactions with industrial hygienist (set of dummy variables representing the 42 industrial hygienists) as the independent variables. A significant training effect ( $p < .0001$ ), significant company effect ( $p = 0.0020$ ), hygienist-company interaction ( $p = 0.0017$ ), and hygienist-training interaction ( $p = 0.0009$ ) was observed again. Since there was a significant improvement observed in the data interpretation test accuracy of industrial hygienists after the data interpretation training, null Hypothesis 1 is rejected.

## Accuracy Results for Task Judgments

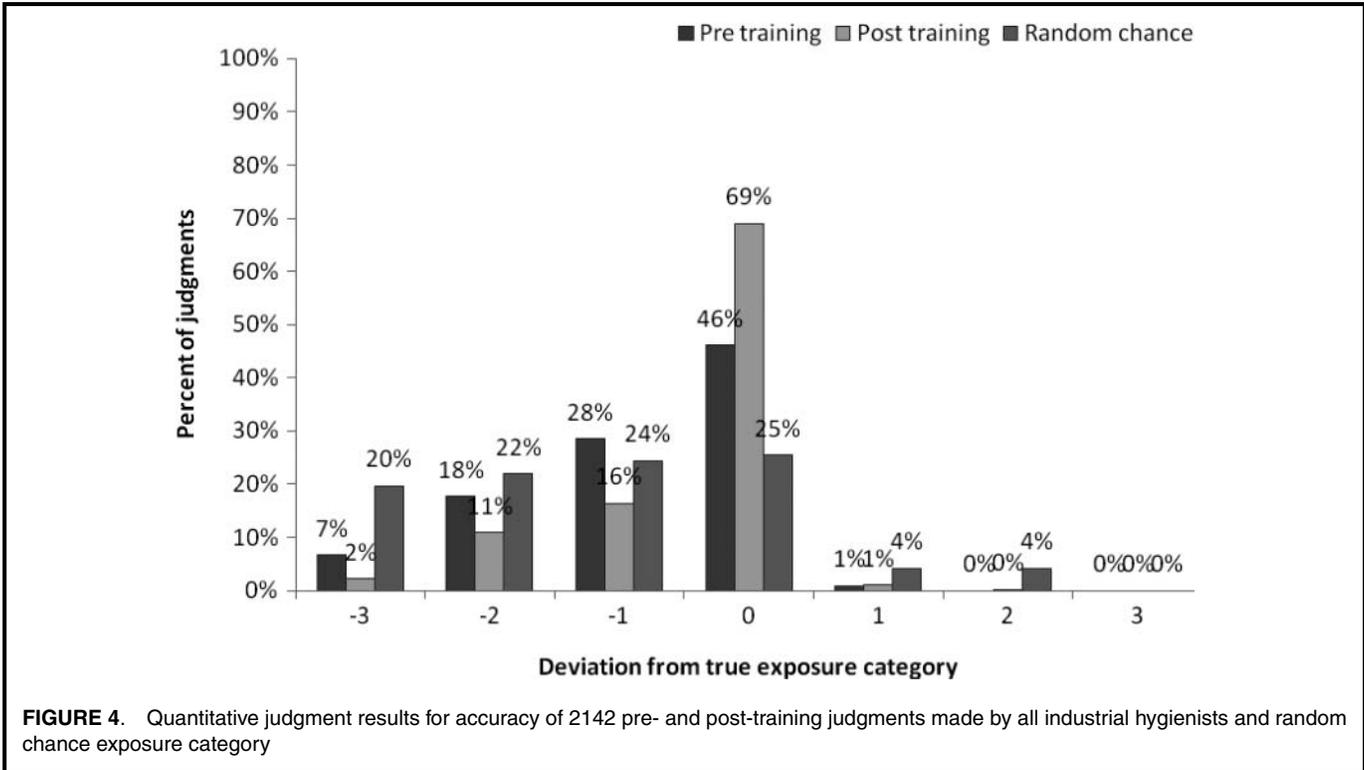
While companies were asked to choose tasks for evaluation that spanned the entire range of AIHA exposure categories, study results indicate that 18 out of 22 tasks ultimately turned out to be Category 4, based on monitoring data. This imbalance limits this study's ability to examine potential bias toward overestimating exposure categories.

Individual quantitative and qualitative judgment categories were compared to the reference category to study accuracy (Figures 4 and 5). For quantitative judgments, the underprediction rate decreased from 53% pre-training to 29% post-training. The overprediction rates, however, remained constant at 1% both pre- and post-training. When compared to judgments made completely randomly (25% accuracy), industrial hygienists did better and predicted the true exposure category 46% of the time ( $p < 0.0001$ ) before training. Significant improvement was observed post-training with the percent of correct judgments increasing from 46% to 69% ( $p < 0.0001$ ).

Qualitative judgments are made without the aid of monitoring data and represent those judgments made only with basic characterization information. Data interpretation training is not expected to have any effect on the qualitative judgments. In any case, post-training qualitative judgments cannot be examined since monitoring data revealed to the industrial hygienists in the pre-training phase would undoubtedly affect post-training qualitative judgments. Therefore, only pre-training qualitative judgment accuracy results are presented (Figure 5), and these are compared to judgments made purely randomly. Only 35% of judgments made by industrial hygienists correctly predicted the correct exposure category pre-training compared to random chance that would predict 25% correct judgments ( $p = 0.07$ ). Sixty-five percent of the actual judgments underpredicted the exposure by 1, 2, or 3 categories compared to 67% predicted by random chance. Thus, the hygienists' accuracy is not significantly different from random chance ( $p = 0.078$ ).

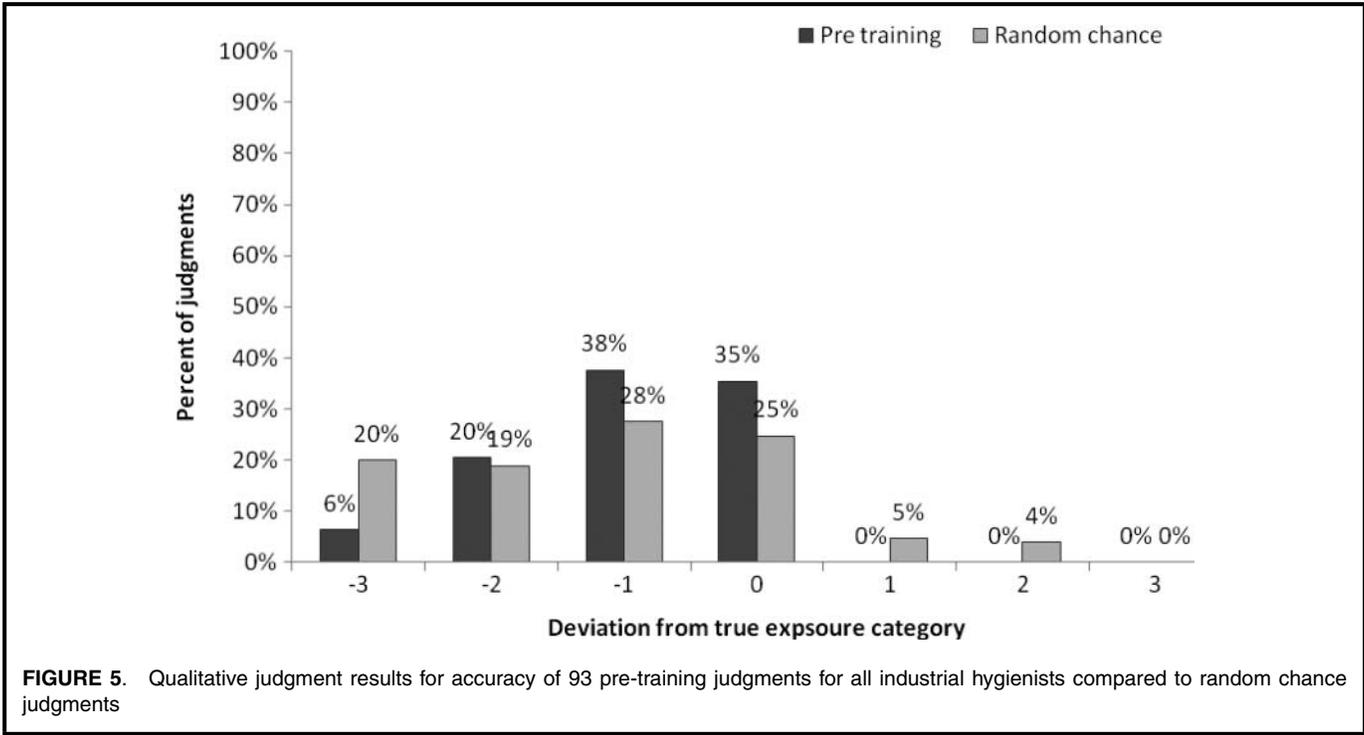
Again, because of the limitation of the data being heavily weighted to Category 4 exposures, overestimation bias could not be studied. It is not within the scope of this article to discuss how to improve qualitative judgments.

The fraction of correct decisions made by each industrial hygienist was investigated pre- and post-training (Table AI in the Appendix). It was found that the rule of thumb training improved the decisions made by most industrial hygienists. For example, Hygienists 1, 5, 9, 26, 27, 29, 37, 43, and 44 had zero correct decisions pre-training, but the fraction of correct decisions made by these hygienists post-training was 0.47, 0.45, 0.28, 0.06, 1.00, 0.44, 0.33, 0.11, and 0.89, respectively (Table AII in the Appendix). Table VI shows the summary results of the fraction of correct judgments made by all 49 industrial hygienists. Although four of the hygienists seemed to get worse after training, an overall improvement was observed, with 82% of hygienists making correct judgments after training, indicating a positive training effect. Ten percent of hygienists were not able to predict the correct judgment before training and did not change their judgment even after training.



We can study the fraction of hygienists making correct quantitative judgments after each new monitoring data point has been revealed. Since there are 161 monitoring data points for 17 tasks, this allows us 161 occasions (pre- and post-training) to study the effect of training. A marked increase was

observed post-training in the fraction of hygienists making correct judgments. Table AII in the Appendix shows the results for fraction of hygienists making correct decisions at each data point for each task, and these are summarized in Table VII. It can be seen that out of the 161 occasions, hygienists improved



**TABLE VI. Effect of Training on Judgment Accuracy for All 49 Industrial Hygienists**

| Pre- and Post-Training Judgment Accuracy      | No. of Hygienists (%) |
|---|-----------------------|
| Correct judgment before and after training    | 7 (14)                |
| Improvement in judgments after training       | 33 (68)               |
| Deterioration in judgments after training     | 4 (8)                 |
| Incorrect judgments before and after training | 5 (10)                |
| Total   | 49 (100)              |

90 times post-training. Industrial hygienists could not predict the correct exposure category 30 times even after training. Twenty-five percent of hygienists were able to predict the correct exposure category before training, and their judgment stayed the same even after training. It is evident that the focused training on assessing the 95th percentile helps industrial hygienists make better judgments. This can be seen in both the improvement in DIT judgments as well as improvement in the quantitative judgments for different tasks.

To test Hypothesis 2 (effect of training on accuracy), absolute accuracy and categorical accuracy of task judgments were modeled using SAS. For absolute accuracy, a general linear model was fitted with absolute accuracy as the dependent variable and company (set of dummy variables representing the six companies), training (binary variable, 0 = pre-training and 1 = post-training), number of data points (continuous variable, 1 through 24), hygienist interactions with company, hygienist interaction with training, and company and task interactions as independent variables (Table VIII). For categorical accuracy, a cumulative logit model was fitted with categorical accuracy as the dependent variable and company (set of dummy variables representing the six companies), training (binary variable, 0 = pre-training and 1 = post-training), number of data points (continuous variable, 1 through 24), and their interactions as independent variables.

The study design allowed us to choose the company and training as the independent variables and since there were varying amounts of data available for each task, the number of data point available for a given judgment was also selected as a variable to be included in the accuracy models. The absolute accuracy model was found to be significant ( $p < 0.05$ ), with an  $r$  value of 0.66. A significant training effect was found with absolute accuracy ( $p < 0.0001$ ) and also with categorical accuracy ( $p < 0.0001$ ). This indicated that post-training, industrial hygienists were able to predict the correct exposure category more accurately.

Company was found to be a significant predictor of accuracy, indicating that industrial hygienists from certain companies did better. The number of monitoring data used in making a judgment was found to be significant ( $p < .0001$ ) for both task absolute accuracy and task categorical accuracy. Thus, when more data points were revealed, the industrial hygienists

**TABLE VII. Fraction of Industrial Hygienists Making Correct Decisions for 161 Task-Data Point Scenarios**

| Pre- and Post-Training Judgment Accuracy of Hygienists                     | No. Task-Data Point Scenarios (%) |
|--|-----------------------------------|
| All hygienists making correct judgments before and after training          | 40 (25)                           |
| Increase in fraction of hygienists making correct judgments after training | 90 (56)                           |
| Decrease in fraction of hygienists making correct judgments after training | 1 (0.6)                           |
| All hygienists making incorrect judgments before and after training        | 30 (18.4)                         |
| Total  | 161 (100)                         |

were better able to predict the correct category. Results for categorical accuracy were similar to the results for absolute accuracy. Hence, null Hypothesis 2 is rejected. There is a significant improvement in accuracy after training.

Significant effects were seen for company, the interaction of company and task, and the interactions of company and industrial hygienist, and training and industrial hygienist for absolute accuracy and categorical accuracy. This shift toward improvement in correct exposure category prediction is also evident from Figure 4. The interaction of industrial hygienist and company was found to be significant, indicating that there was not only an overall company effect, but even within a company, certain industrial hygienists were better at accurately predicting the correct exposure category than the rest. A significant hygienist interaction with training was observed, indicating that, again, certain industrial hygienists are better at accurately predicting the correct exposure category post-training even though all the industrial hygienists were provided with the same training. These hygienist interactions with company and training suggest the need to further investigate the characteristics of the industrial hygienists that might influence accuracy. The company task interaction was found to be significant, indicating that accuracy of industrial hygienist from a company varied by the task they evaluated.

**TABLE VIII. Results from General Linear Model (GLM) with Absolute Accuracy as the Dependent Variable**

| Variable           | F value | Pr > F |
|--------------------|---------|--------|
| Company            | 46.73   | <.0001 |
| Training           | 303.89  | <.0001 |
| Data point         | 117.42  | <.0001 |
| Hygienist*Company  | 18.51   | <.0001 |
| Hygienist*Training | 5.3     | <.0001 |
| Company*Task       | 103.85  | <.0001 |

## CONCLUSIONS

It is important for professional industrial hygienists to be able to accurately predict the correct exposure category based on limited data because exposure control decisions are driven by these judgments. This study hypothesized that data interpretation training would help improve accuracy. The results of the analyses of absolute and categorical accuracy for both data interpretation tests and quantitative judgments for actual tasks show that targeted training in the interpretation of lognormal data does improve industrial hygienists' judgment accuracy.

There was substantial underestimation bias seen in the pre-training data interpretation test results. Post-training, this underestimation bias was reduced or shifted slightly toward overestimation. The percentage of underestimation of exposures decreased for quantitative judgments after training, although the high percentage of Category 4 exposure tasks in the study make it difficult to evaluate any overestimation bias effect. Hygienists on average performed better than random chance for the data interpretation tests as well as quantitative judgments. This "expertise" in judging exposures could be due to their training and experience. These determinants of accuracy need to be investigated further.

If a limited amount of training based on simple rules of thumb can significantly improve judgment accuracy, routine use of standard statistical tools will increase accuracy to an even greater extent and eliminate the underestimation bias altogether. When characterizing exposures and making decisions regarding exposure controls, approaches that are highly accurate or that slightly overestimate exposures are preferred over approaches that underestimate exposures.

The amount of monitoring data was also found to have a significant effect on improving both absolute and categorical accuracy of quantitative judgments. Significant effects were seen by company and by the interaction of company and hygienist. Company was found to be a significant variable, indicating that there may be some characteristics of a company that affect accurate judgments or that it could be a surrogate for certain qualities possessed by hygienists within that company. For example, some companies may hire only CIHs, or a large percentage of their hygienists are CIHs, which may, in turn, affect judgment accuracy. There may also be company-specific factors, e.g., some companies may have a well-defined exposure assessment strategy akin to the AIHA strategy. Therefore, industrial hygienists in that company may already be familiar with the strategy and do better in this study. Certain companies could have a better training program in place and may even be using statistical tools for day-to-day exposure assessment.

Other causes for the apparent effect can also be postulated. For example, the potential company effect may be related to the specific tasks evaluated within each company and that being trained on specific tasks for a long period of time could also help improve accuracy. The design of this study limits its ability to evaluate the various hypotheses that can be postulated to explain the company effect. Hygienist interactions were

found to be significant, indicating that there may be certain qualities of hygienists such as years of experience that make them more accurate. Hence, hygienist-specific information like education and experience determinants will be examined more closely in forthcoming papers.

The results of this study should be interpreted with caution. The study sample size gave reasonable statistical power for most of the results (number of industrial hygienists = 49 and number of scenarios = 2235), and there was a good balance of experience level among study participants that varied from 4 months to 26 years. However, although the tasks were from different companies, they are not representative of all tasks across industry and were heavily weighted toward Category 4 exposures.

The accuracy of qualitative judgments was low and not significantly better than predicted by random chance. Improvement in qualitative judgments was not studied due to the nature of the study design. For improving qualitative judgments, options such as training in the use of exposure models and rules of thumb for evaluating basic characterization information should be explored in future studies.

Results show significant improvements in the accuracy of quantitative judgments of hygienists as a result of training. Hence, the use of statistical tools and development of standard training programs is recommended to improve judgments based on monitoring data and provide feedback and calibration to improve qualitative judgments.

## ACKNOWLEDGMENTS

This project was supported by grant 1RO1 OH008513 from the National Institute for Occupational Safety and Health of the Centers for Disease Control and Prevention.

## REFERENCES

1. **Corn, M., and N.E. Esmen:** Workplace exposure zones for classification of employee exposures to physical and chemical agents. *Am. Ind. Hyg. Assoc. J.* 40:47–57 (1979).
2. **Mulhausen, J.R., and J. Damiano:** *A Strategy for Assessing and Managing Occupational Exposures*, Second Ed. Fairfax, Va.: AIHA Press, 1998.
3. **Ignacio, J., and B. Bullock (eds.):** *A Strategy for Assessing and Managing Occupational Exposures*, Third Ed. Fairfax, Va.: AIHA Press, 2006.
4. **Hawkins, N.C., and J.S. Evans:** Subjective estimation of toluene exposures: A calibration study of industrial hygienists. *Appl. Ind. Hyg.* 4:61–68 (1989).
5. **Walker, K.D., D. Macintosh, and J.S. Evans:** Use of expert judgment in exposure assessment, Part I: Characterization of personal exposures to benzene. *J. Expo. Anal. Environ. Epidemiol.* 11:308–322 (2001).
6. **Walker, K.D., P. Catalano, J.S. Evans, and J.K. Hammitt:** Use of expert judgment in exposure assessment, part II: Calibration of expert judgments about personal exposures to benzene. *J. Expo. Anal. Environ. Epidemiol.* 13:1–16 (2003).
7. **Logan, L., G. Ramachandran, J. Mulhausen, and P. Hewett:** Occupational exposure decisions: Can limited data interpretation training help improve accuracy? *Ann. Occup. Hyg.* 53(4):311–324 (2009).

8. **Macaluso, M., E. Delzell, V. Rose, J. Perkins, and K. Oestenstad:** Inter-rater agreement in the assessment of solvent exposure at a car assembly plant. *Am. Ind. Hyg. Assoc. J.* 54:351–359 (1993).
9. **DeCock, J., H. Kromhout, D. Heederik, and J. Burema:** Experts' subjective assessment of pesticide exposure in fruit growing. *Scand. J. Work Environ. Health* 22(6):425–432 (1996).
10. **Kromhout, H., Y. Oostendorp, D. Heederik, and J.S.M. Boleij:** Agreement between semiquantitative exposure estimates and quantitative exposure measurements. *Am. J. Ind. Med.* 12:551–562 (1987).
11. **Stewart, P.A., R. Caret, C. Schairer, and A. Blair:** Comparison of industrial hygienists' exposure evaluations for an epidemiologic study. *Scand. J. Work Environ. Health* 26:44–51 (2000).
12. **Post, W., H. Kromhout, D. Heederik, D. Noy, and R.S. Duijzentkunst:** Semiquantitative estimates of exposure to methylene chloride and styrene: The influence of quantitative exposure data. *Appl. Occup. Environ. Hyg.* 6:197–204 (1991).
13. **Kauppinen, T., K. Teschke, G. Astrakianakis, et al.:** Assessment of exposure in an international study on cancer risks among pulp, paper, and paper product workers. *Am. Ind. Hyg. Assoc. J.* 63:254–261 (2002).
14. **Goldberg, M.S., J. Siemiatycki, and M. Gerin:** Inter-rater agreement in assessing occupational exposures in a case-control study. *Br. J. Ind. Med.* 43:667–676 (1986).
15. **Benke, G., M. Sim, A. Forbes, and M. Salzberg:** Retrospective assessment of occupational exposure to chemicals in community-based studies: Validity and repeatability of industrial hygiene panel rating. *Int. J. Epidemiol.* 26:635–642 (1997).
16. **Rybicki, B.A., C.C. Johnson, E.L. Peterson, G.X. Kortsha, and J.M. Gorell:** Comparability of different methods of retrospective exposure assessment of metals in manufacturing industries. *Am. J. Ind. Med.* 31:36–43 (1997).
17. **Rybicki, B.A., E.L. Peterson, C.C. Johnson, G.X. Kortsha, W.M. Cleary, and J.M. Gorell:** Intra- and inter-rater agreement in assessment of occupational exposure to metals. *Int. J. Epidemiol.* 27:269–273 (1998).
18. **Hewett, P., P. Logan, J. Mulhausen, G. Ramachandran, and S. Banerjee:** Rating exposure control using Bayesian decision analysis. *J. Occup. Environ. Hyg.* 3:568–581 (2006).
19. **Agresti, A.:** *Categorical Data Analysis*, Second Ed. Hoboken, N.J.: John Wiley & Sons, Inc., 2002.

**APPENDIX**

**TABLE A1. Fraction of Correct Decisions Made by Hygienists, and the Difference in the Judgments Before and After Training, Showing an Overall Improvement Trend**

| Hygienist | Fraction of Correct Decision<br>Pre-Training | Fraction of Correct Decisions<br>Post-Training | Difference<br>(Post – Pre) |
|-----------|--|--|----------------------------|
| 1         | 0.00   | 0.47   | 0.47                       |
| 2         | 0.33   | 0.80   | 0.47                       |
| 3         | 0.25   | 0.13   | -0.13                      |
| 4         | 0.69   | 0.97   | 0.28                       |
| 5         | 0.00   | 0.45   | 0.45                       |
| 6         | 0.52   | 0.90   | 0.38                       |
| 7         | 0.07   | 0.93   | 0.86                       |
| 8         | 0.10   | 0.48   | 0.38                       |
| 9         | 0.00   | 0.28   | 0.28                       |
| 10        | 0.31   | 0.90   | 0.59                       |
| 11        | 0.24   | 0.41   | 0.17                       |
| 12        | 0.21   | 0.83   | 0.62                       |
| 13        | 1.00   | 0.62   | -0.38                      |
| 14        | 0.52   | 0.61   | 0.09                       |
| 15        | 0.61   | 0.61   | 0.00                       |
| 16        | 0.57   | 0.61   | 0.04                       |
| 17        | 0.65   | 0.65   | 0.00                       |
| 18        | 0.26   | 0.61   | 0.35                       |
| 19        | 0.61   | 0.61   | 0.00                       |
| 20        | 0.57   | 0.61   | 0.04                       |
| 21        | 0.61   | 0.69   | 0.07                       |
| 22        | 0.51   | 0.69   | 0.17                       |
| 23        | 0.56   | 0.70   | 0.14                       |
| 24        | 0.50   | 0.83   | 0.33                       |
| 25        | 0.59   | 0.82   | 0.24                       |
| 26        | 0.00   | 0.06   | 0.06                       |
| 27        | 0.00   | 1.00   | 1.00                       |
| 28        | 0.13   | 1.00   | 0.88                       |
| 29        | 0.00   | 0.44   | 0.44                       |
| 30        | 1.00   | 1.00   | 0.00                       |
| 31        | 0.78   | 0.67   | -0.11                      |
| 32        | 1.00   | 1.00   | 0.00                       |
| 33        | 0.89   | 1.00   | 0.11                       |
| 34        | 1.00   | 1.00   | 0.00                       |
| 35        | 0.56   | 0.78   | 0.22                       |
| 36        | 0.67   | 0.89   | 0.22                       |
| 37        | 0.00   | 0.33   | 0.33                       |
| 38        | 0.67   | 0.78   | 0.11                       |
| 39        | 0.89   | 0.67   | -0.22                      |
| 40        | 0.89   | 0.89   | 0.00                       |
| 41        | 0.56   | 0.89   | 0.33                       |
| 42        | 1.00   | 1.00   | 0.00                       |
| 43        | 0.00   | 0.11   | 0.11                       |
| 44        | 0.00   | 0.89   | 0.89                       |
| 45        | 1.00   | 1.00   | 0.00                       |
| 46        | 1.00   | 1.00   | 0.00                       |
| 47        | 0.89   | 1.00   | 0.11                       |
| 48        | 1.00   | 1.00   | 0.00                       |
| 49        | 0.78   | 0.78   | 0.00                       |

*Note:* Each hygienist evaluated between 5 and 24 scenarios in tasks within his or her company.

**TABLE All. Fraction of Hygienists Making Correct Decisions at Each Data Point for 17 Tasks**

| Data Points | Task 1 Training |      | Task 2 Training |      | Task 3 Training |      | Task 4 Training |      | Task 5 Training |      | Task 6 Training |      |
|-------------|-----------------|------|-----------------|------|-----------------|------|-----------------|------|-----------------|------|-----------------|------|
|             | Pre             | Post |
| 1           | 0.30            | 0.60 | 0.40            | 1.00 | 0.86            | 1.00 | 1.00            | 1.00 | 0.00            | 0.00 | 1.00            | 1.00 |
| 2           | 0.70            | 0.10 | 0.40            | 1.00 | 1.00            | 1.00 | 1.00            | 1.00 | 0.00            | 0.00 | 1.00            | 1.00 |
| 3           | 0.40            | 0.60 | 0.40            | 1.00 | 0.86            | 1.00 | 1.00            | 1.00 | 0.00            | 0.00 | 1.00            | 1.00 |
| 4           | 0.30            | 0.80 | 0.30            | 0.78 | 0.86            | 1.00 | 1.00            | 1.00 | 0.00            | 0.00 | 1.00            | 1.00 |
| 5           | 0.10            | 0.80 | 0.20            | 0.67 | 0.86            | 1.00 | 1.00            | 1.00 | 0.00            | 0.00 | 0.67            | 1.00 |
| 6           | 0.10            | 0.80 | 0.30            | 0.78 | 0.86            | 1.00 |                 |      | 0.00            | 0.00 | 0.67            | 1.00 |
| 7           | 0.60            | 0.90 | 0.30            | 0.78 | 0.86            | 1.00 |                 |      | 0.00            | 0.00 | 1.00            | 1.00 |
| 8           | 0.30            | 0.80 | 0.30            | 0.78 |                 |      |                 |      | 0.00            | 0.00 | 1.00            | 1.00 |
| 9           | 0.50            | 0.90 | 0.30            | 0.78 |                 |      |                 |      |                 |      | 1.00            | 1.00 |
| 10          | 0.30            | 0.70 | 0.10            | 0.78 |                 |      |                 |      |                 |      | 1.00            | 1.00 |
| 11          |                 |      | 0.10            | 0.67 |                 |      |                 |      |                 |      | 1.00            | 1.00 |
| 12          |                 |      | 0.20            | 0.89 |                 |      |                 |      |                 |      | 1.00            | 1.00 |
| 13          |                 |      | 0.30            | 0.67 |                 |      |                 |      |                 |      | 1.00            | 1.00 |
| 14          |                 |      | 0.30            | 0.67 |                 |      |                 |      |                 |      | 1.00            | 1.00 |
| 15          |                 |      | 0.30            | 0.67 |                 |      |                 |      |                 |      | 1.00            | 1.00 |
| 16          |                 |      | 0.30            | 0.67 |                 |      |                 |      |                 |      | 1.00            | 1.00 |
| 17          |                 |      | 0.30            | 0.67 |                 |      |                 |      |                 |      | 1.00            | 1.00 |
| 18          |                 |      |                 |      |                 |      |                 |      |                 |      | 1.00            | 1.00 |
| 19          |                 |      |                 |      |                 |      |                 |      |                 |      | 1.00            | 1.00 |
| 20          |                 |      |                 |      |                 |      |                 |      |                 |      | 1.00            | 1.00 |
| 21          |                 |      |                 |      |                 |      |                 |      |                 |      | 1.00            | 1.00 |
| 22          |                 |      |                 |      |                 |      |                 |      |                 |      | 1.00            | 1.00 |
| 23          |                 |      |                 |      |                 |      |                 |      |                 |      | 1.00            | 1.00 |
| 24          |                 |      |                 |      |                 |      |                 |      |                 |      | 1.00            | 1.00 |

| Data Points | Task 7 Training |      | Task 8 Training |      | Task 9 Training |      | Task 10 Training |      | Task 11 Training |      | Task 12 Training |      |
|-------------|-----------------|------|-----------------|------|-----------------|------|------------------|------|------------------|------|------------------|------|
|             | Pre             | Post | Pre             | Post | Pre             | Post | Pre              | Post | Pre              | Post | Pre              | Post |
| 1           | 1.00            | 1.00 | 1.00            | 1.00 | 0.00            | 0.00 | 0.00             | 0.00 | 0.00             | 0.00 | 0.5              | 1    |
| 2           | 1.00            | 1.00 | 1.00            | 1.00 | 0.00            | 0.00 | 0.00             | 0.00 | 0.00             | 0.00 | 0.5              | 1    |
| 3           | 1.00            | 1.00 | 1.00            | 1.00 | 0.00            | 0.00 | 0.00             | 0.00 | 0.00             | 0.00 | 0.5              | 1    |
| 4           | 1.00            | 1.00 | 1.00            | 1.00 | 0.00            | 0.00 | 0.00             | 0.33 | 0.00             | 0.00 | 0.5              | 1    |
| 5           | 1.00            | 1.00 | 1.00            | 1.00 | 0.00            | 0.00 | 0.33             | 1.00 | 0.00             | 0.00 | 0.5              | 1    |
| 6           | 1.00            | 1.00 | 1.00            | 1.00 | 0.00            | 0.00 | 0.00             | 0.67 | 0.00             | 0.00 | 0.5              | 1    |
| 7           |                 |      |                 |      | 0.00            | 1.00 | 0.00             | 0.67 | 0.00             | 1.00 | 0.5              | 1    |
| 8           |                 |      |                 |      |                 |      | 0.33             | 1.00 | 0.00             | 0.67 | 0.5              | 1    |
| 9           |                 |      |                 |      |                 |      | 0.67             | 1.00 | 0.00             | 0.67 |                  |      |
| 10          |                 |      |                 |      |                 |      | 0.67             | 1.00 | 0.00             | 1.00 |                  |      |
| 11          |                 |      |                 |      |                 |      | 0.67             | 1.00 |                  |      |                  |      |

| Data Points | Task 13 Training |      | Task 14 Training |      | Task 15 Training |      | Task 16 Training |      | Task 17 Training |      |
|-------------|------------------|------|------------------|------|------------------|------|------------------|------|------------------|------|
|             | Pre              | Post |
| 1           | 0                | 0.5  | 0                | 0.5  | 0.00             | 0.5  | 0.50             | 0.5  | 0.79             | 0.79 |
| 2           | 0                | 0.5  | 0                | 0.5  | 0.00             | 0.5  | 0.00             | 1    | 0.84             | 0.84 |
| 3           | 0                | 0.5  | 0                | 0    | 0.00             | 1    | 0.00             | 1    | 0.68             | 0.79 |
| 4           | 0.5              | 1    | 0                | 0    | 0.00             | 0.5  | 0.00             | 1    | 0.47             | 0.89 |
| 5           | 0                | 1    | 0                | 0.5  | 0.00             | 0.5  | 0.00             | 1    | 0.79             | 0.89 |

**TABLE AII. (Continued)**

| Data Points | Task 13 Training |      | Task 14 Training |      | Task 15 Training |      | Task 16 Training |      | Task 17 Training |      |
|-------------|------------------|------|------------------|------|------------------|------|------------------|------|------------------|------|
|             | Pre              | Post |
| 6           | 0                | 1    | 0                | 0    | 0.00             | 0.5  | 0.00             | 1    | 0.84             | 1.00 |
| 7           | 0                | 1    | 0                | 0.5  | 0.00             | 0.5  | 0.50             | 1    | 0.79             | 0.89 |
| 8           |                  |      | 0                | 0    | 0                | 0.5  | 0.5              | 1    | 0.84             | 0.95 |
| 9           |                  |      |                  |      |                  |      | 0.5              | 1    |                  |      |
| 10          |                  |      |                  |      |                  |      | 0.5              | 1    |                  |      |
| 11          |                  |      |                  |      |                  |      | 0.5              | 1    |                  |      |